Cognitive Fit Between Conceptual Schemas and Internal Problem Representations: The Case of Geospatio–Temporal Conceptual Schema Comprehension

—VIJAY KHATRI, MEMBER, IEEE, IRIS VESSEY, SUDHA RAM, MEMBER, IEEE, AND V. RAMESH

Abstract—*Geospatio–temporal conceptual models provide a mechanism to explicitly represent geospatial and temporal aspects of applications. Such models, which focus on both "what" and "when/where," need to be more expressive than conventional conceptual models (e.g., the ER model), which primarily focus on "what" is important for a given application. In this study, we view conceptual schema comprehension of geospatio–temporal data semantics in terms of matching the external problem representation (that is, the conceptual schema) to the problem-solving task (that is, syntactic and semantic comprehension tasks), an argument based on the theory of cognitive fit. Our theory suggests that an external problem representation that matches the problem solver's internal task representation will enhance performance, for example, in comprehending such schemas. To assess performance on geospatio–temporal schema comprehension tasks, we conducted a laboratory experiment using two semantically identical conceptual schemas, one of which mapped closely to the internal task representation while the other did not. As expected, we found that the geospatio–temporal conceptual schema that corresponded to the internal representation of the task enhanced the accuracy of schema comprehension; comprehension time was equivalent for both. Cognitive fit between the internal representation of the task and conceptual schemas with geospatio–temporal annotations was, therefore, manifested in accuracy of schema comprehension and not in time for problem solution. Our findings suggest that the annotated schemas facilitate understanding of data semantics represented on the schema.*

Index Terms—*Conceptual modeling, geospatial database, geospatio–temporal conceptual models, human associative memory (HAM), syntactic and semantic comprehension tasks, temporal database, theory of cognitive fit.*

Most applications require some aspect of time in organizing their information, for example, healthcare (patient histories), insurance (claims and accident histories), reservation systems, and scientific applications. Many applications also require some aspect of space; according to Albaredes, 80% of all human decisions contain a spatial component [1]. In research on perception, space is differentiated into LARGE-SCALE and SMALL-SCALE space [2]. The former, which is also referred to as geographic/geospatial space, is defined as one that cannot be viewed from a single vantage point, while the latter is visible from a single vantage point. In this research, we focus on the former.

Manuscript received November 9, 2004; revised November 14, 2005.
V. Khatri and V. Ramesh are with the Kelley School of Business, Indiana University, Bloomington, IN 47405 USA (email: vkhatri@indiana.edu; venkat@indiana.edu).
I. Vessey is with the University of Queensland and Queensland University of Technology, Brisbane, QLD 4000, Australia (email: i.vessey@business.uq.edu.au).
S. Ram is with the Eller College of Management, University of Arizona, Tucson, AZ 85721 USA (email: ram@eller.arizona.edu).

IEEE DOI 10.1109/TPC.2006.875091

Mennecke and Crossland describe how geospatial information can be applied to business applications such as facility management, market analysis, transportation, logistics, strategic planning, and decision-making [3]. Retailers such as Ace Hardware Corporation have used geographic information to identify underserved customers and make decisions related to store relocation based on the location of their customers [4]. Underlying the applications described above are temporal and geospatial data, collectively referred to as GEOSPATIO–TEMPORAL data.

In developing geospatio–temporal applications, there is a need to elicit the data semantics not only related to **what** is important for the application but also related to **when** and **where**. One of the problems with designing such applications is that there is "a gulf between the richness of knowledge structures in application domains and the relative simplicity of the data model in which the structures can be expressed" [5]. **Conventional conceptual models** (see, for example, [6]–[8]) that provide a formalism to represent what is pertinent for an application are, at best, only partially useful for geospatio–temporal applications (see, for example, [9]). To help represent the geospatio–temporal data requirements, a number of modeling formalisms have been proposed. For examples of formalisms that support explicit representation of temporal data semantics, see [10], and for geospatial data semantics, see [11] and [12].

Shoval and Frumermann assert that a conceptual model should have "semantic expressiveness," where expressiveness refers to the availability of a large variety of concepts for a more comprehensive representation of the real world [13]. Since conceptual schemas drive discovery, they should also be clear and comprehensible. The primary challenge in representing when/where along with what is balancing simplicity and understandability with expressiveness [14]. Hence, it is important to understand the tradeoff between expressiveness and understandability in geospatio–temporal conceptual schema comprehension.

In this study, we address the research question: How should geospatio–temporal data semantics be represented to effectively support data analysts in schema comprehension? Two major theoretical perspectives dominate the literature on human factors: that addressing the cognitive characteristics of displays is exemplified by the proximity compatibility principle [15]; that addressing the cognitive characteristics of both displays and the task to which they are being applied is exemplified by the theory of cognitive fit. Because the theory of cognitive fit is more specific to the task at hand, it represents a stronger theoretical approach than does the proximity compatibility principle. (See Vessey and Glass [16] and Newell [17] for the fundamental concepts relating to weak and strong approaches to problem solving.) Prior research (see, for example, [18]) suggests that conceptual models for geographic data representation do not explicitly incorporate how humans cognitively store and use geographic data. Hence, our objective in this study is to develop a theoretical explanation for the effectiveness of representations in facilitating understanding of geospatio–temporal data semantics—that is, a specific task context—that is based on research in cognition. To do so, we elaborate on a further aspect of the theory of cognitive fit. We address the applicability of the proximity compatibility principle to the study we conducted in the discussion section of the pape

The rest of the paper is organized as follows. In the next section, we examine background research related to geospatio–temporal conceptual modeling. In the following section, we present the theory that forms the foundation for our investigation. We then present the methodology used to test the propositions, followed by the data analysis. The paper concludes with implications for research and practice.

PRIOR RESEARCH

In this section, we briefly describe the characteristics of the geospatio–temporal data semantics that are pertinent for the representation of requirements for geospatial and temporal applications. We then describe various formalisms that can be used to represent these requirements. Finally, we review prior research on conceptual schema understanding tasks.

Characterizing Geospatio–Temporal Data Semantics Wand, Monarchi, Parsons, and Woo suggest that an ontology be employed to define concepts in a modeling language [19]. The basis of a geospatio–temporal conceptual model is a time and space ontology that defines concepts like event and state [20], valid time and transaction time [21], lifespan (or existence time) [22], temporal and geospatial granularities along with indeterminacy [23], and geometry and position [24]. We provide brief definitions of these terms below; for more details on geospatio–temporal ontology, the reader is referred to [14] and [23].

An EVENT occurs at a point in time, that is, it has no duration (for example, lightning hit the road at 2:03 P.M.), while a STATE has duration (for example, a storm occurred from 5:07 P.M. to 5:46 P.M.). While VALID TIME denotes when the fact is true in the real world and implies the storage of histories related to facts, TRANSACTION TIME links an object to the time it is current in the database and implies the storage of versions of a database object. The data semantics of valid time associated with a fact imply that the fact can exist at certain points in time (events) or in certain time periods (states) in the past, the present, or the future. On the other hand, the data semantics of transaction time associated with an object require that the object can exist in certain time periods in the past until now (state). EXISTENCE TIME, which applies to an object, is the valid time when the object exists. POSITION in space is based on coordinates in a mathematically defined reference system, for example, latitude and longitude. The shape of the object is represented by GEOMETRY, for example, point, line, and region. GRANULARITIES are intrinsic to geospatial and temporal data, and provide a mechanism to hide details that are not known or not pertinent for an application. For example, in a cadastral application [25], mortgages can be associated with a temporal granularity of *day* and the representation of long-term land-use changes may require a temporal granularity of *year*. *Day* and *year*, or more accurately *Gregorian day* and *Gregorian year*, are examples of temporal granularities, which belong to the Gregorian calendar.

Representing Geospatio–Temporal Data Semantics The first generation of geospatio–temporal conceptual models (see, for example, [26] and [27]) was map- or GIS-oriented. Certain of the next-generation geospatio–temporal conceptual models represented the additional aspects of the application by changing the semantics of conventional conceptual models without adding any new constructs (see, for example, [28] and [29]), while others added new constructs (see, for example, [30]). For example, the Temporal EER (TEER) Model assumes that **all** entities have a lifespan and that **all** attributes are temporal [28], [29]; the syntax in a conventional conceptual model, for example, a rectangle used to represent an entity type, was therefore ascribed a new meaning. As a further example, the Relationship, Attributes, Keys, and Entities (RAKE) Model includes **new** constructs for temporal aspects like true events, or “durationless states” [30, p. 285].

On the other hand, certain authors (see, for example, [14] and [31]) argue that, in order to simplify the complex task of representing geospatio–temporal data semantics, geospatio–temporal aspects should be the *last* consideration in conceptual design (see, Fig. 1). As shown in Fig. 1, there are two levels of abstraction, one for what and the other for when/where; an abstraction provides a mechanism for focusing on selected details while deliberately deferring others. While conventional conceptual models can be employed to represent what, when/where can be represented using the supplementary level of abstraction, which is typically provided via geospatio–temporal annotations (see, for example, [12], [14], [22], and [32]). Snodgrass suggests two ways of implementing annotations: (1) presenting the (additional when/where) data semantics on the schema; and (2) listing them separately in text form in the geospatio–temporal data dictionary [31]. While the latter technique is usually employed in practice, many recent research approaches advocate the former.

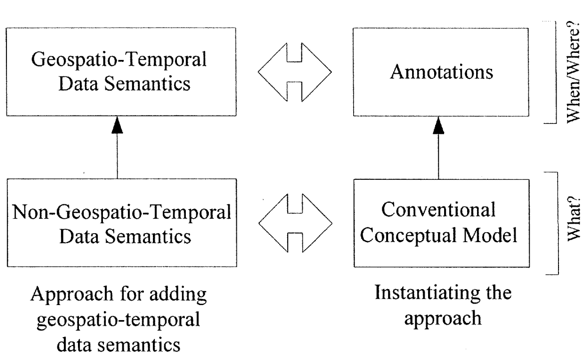

Fig. 1. Representing geospatio–temporal data semantics.

A schema with geospatio–temporal annotations is referred to as a 2-LA (levels of abstraction) schema because it includes two levels of abstraction, one that represents what and the second that represents when/where **on** the schema (see, for example, [12], [14], [22], and [32]). On the other hand, a 1-LA schema, which represents the geospatio–temporal data semantics in the data dictionary includes one level of abstraction on the schema, that related to what (see, for example, [31]). Snodgrass [31], for example, represents the temporal aspects in a table outside the schema. For a valid time relationship, for example, a when/where table includes the name of the relationship, valid-time granularity, and comments. Note that Snodgrass proposes temporal annotations only and that we extended Snodgrass' approach for representing temporal data semantics to also represent geospatial data semantics.

Formalisms that employ 2-LA represent current thinking in this area. To illustrate an annotation-based conceptual model that represents geospatio–temporal data semantics *on* the schema, we present examples using three different formalisms: STER [12]; MADS [32]; and ST USM [14] (see Fig. 2). In a conventional conceptual model, a rectangle is used to represent **entity types** [7]. Fig. 2, for example, denotes that *LAND PARCEL* is an entity type that is pertinent to a database application. In this example, *LAND PARCEL* is represented as a region and has an associated lifespan (or existence time). Each of the three formalisms in Fig. 2 employs different annotation syntax to capture the geospatio–temporal data semantics. STER uses R (region) and *et* (existence time). MADS uses graphical symbols to denote region (region pictogram) and existence time (clock pictogram). ST USM employs a textual string to denote existence time, which is represented as state (S); in Fig. 2, the time periods have a temporal granularity of *day*. Additionally, *LAND PARCEL* is represented as a region (R) in the horizontal plane with a geospatial granularity of *deg* (i.e., degree). Note that the ST USM formalism, which uses a textual string, is more **ontologically expressive** than the alternative formalisms employing annotations because it includes specification of both temporal

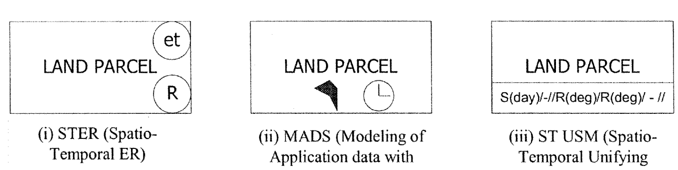

Fig. 2. Capturing the geospatio–temporal semantics using annotations.

and geospatial granularities, while the other two formalisms do not [33].

In summary, annotations provide a succinct way of representing the geospatio–temporal aspects that are important for temporal and geospatial applications. Note that STER, MADS, and ST USM are all based on the ER Model or an extension of the ER Model [6]. Similar to the entity type, other types of constructs (e.g., attribute, subclass, aggregate class) can also be annotated. Khatri, Ram, and Snodgrass present the explicit data semantics associated with these geospatio–temporal annotations [14]. However, it is important to note that we do not know the representational effects [34], that is, the behavioral outcomes, of geospatio–temporal annotations **on** the schema.

Conceptual Schema Understanding In empirical research related to conceptual modeling, researchers have used two broad categories of tasks: construction of schemas and understanding of schemas [35]. While much of the early research in conceptual modeling focused on the schema construction task, considerable attention has been paid more recently to the understanding task (see, for example, [36]). In the current study, we focus on the latter.

One distinction in schema understanding tasks appearing in the literature has been whether participants have access to the schema when performing the tasks. This distinction leads to READ-TO-DO (with access to schema) and READ-TO-RECALL TASKS (without access to the schema) [37]; perhaps the best known read-to-recall studies have been conducted by Weber (see, for example, [36]). We employed read-to-do tasks in our research because they are closer to the real-world situation experienced by professional conceptual designers.

Understanding tasks can be further categorized by focusing on the nature of cognitive processing required to perform the task. Prior to 1999, the most common method for assessing conceptual schema understanding required problem solvers to address a series of tasks, now called COMPREHENSION TASKS (see, for example, [38]), which required participants to answer questions based on modeling constructs. Recent research (see, for example, [36], [39], and [40]) has made the case for using tasks that require a greater level of understanding than comprehension tasks; these tasks are referred to as PROBLEM-SOLVING TASKS. As a preliminary investigation into the efficacy of different representations of recording the geospatio–temporal data semantics, we compared performance with 1-LA and 2-LA on comprehension tasks.

THEORY

In this section, we introduce the theory of cognitive fit [41] as the theoretical basis for suggesting which of the approaches to annotating a conceptual model with geospatio–temporal data semantics is more appropriate for supporting comprehension of conceptual schemas. Anderson and Bower's model of memory serves as the basis for understanding the specific type of fit that applies to our examination of two types of representations of the geospatio–temporal data semantics (that is, 1-LA and 2-LA). We then apply the theory of cognitive fit to the types of tasks and conceptual schema addressed in this research. Finally, we present the hypotheses examined.

Theory of Cognitive Fit Because the theory of cognitive fit was developed explicitly to explain which problem representations are best used to support certain types of tasks [41], it is an appropriate theoretical base for the current research; see Fig. 3.

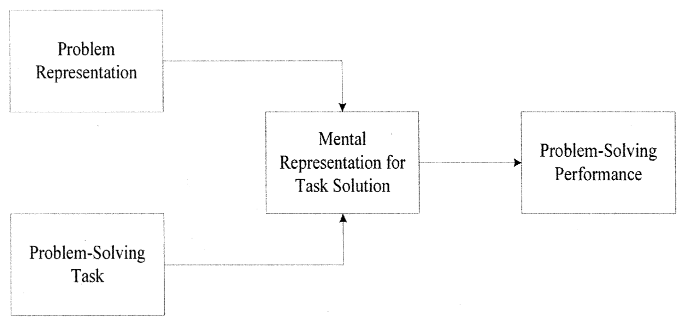

Fig. 3. Cognitive fit in problem solving.

The theory of cognitive fit states that performance (both accuracy and time) on a task will be enhanced when there is a cognitive fit (match) between the information emphasized in the type of problem representation used and that required by the type of problem-solving task under consideration. When the types of information emphasized in the problem-solving elements (in this case problem representation and problem-solving task) match, the problem solver uses processes (and, therefore, formulates a mental representation for task solution) that also emphasizes the same type of information. Consequently, the processes the problem solver uses to both act on the problem representation and to complete the task (problem-solving task) will match, and the problem-solving process will be facilitated. In other words, matching representation to task leads to the use of similar, and therefore, consistent problem-solving processes, and to the formulation of a consistent mental representation for task solution. The problem solver effectively will be guided in solving the problem by the way in which the data is presented and there will be no need to transform the mental representation for task solution to accommodate the use of different processes to extract information from the problem representation and to solve the problem (problem-solving task). Hence, problem solving with cognitive fit leads to effective and efficient problem-solving performance.

The model presented in Fig. 4 presents a recent variant of the basic model of cognitive fit, which has been modified according to Zhang and Norman to incorporate knowledge stored in memory [42] (see, also, [43] and [44]). Zhang and Norman suggest that a cognitive task be viewed as a system of distributed representations with internal and external representations as two indispensable parts [42]. Internal representations are the knowledge structures in the problem solvers' heads, that is, those that can be retrieved from memory—for example, a set of symbols to accomplish a particular task, the rules that govern the use of those symbols, the processes for acting on them, etc. External representations, on the other hand, are the knowledge and structures in the environment—for example, physical symbols, objects, or dimensions, and rules, constraints, or relations embedded in physical configurations [34]. For example, problem solvers use external representations when they use a list for grocery shopping or when they use graphs to understand economic trends [42].

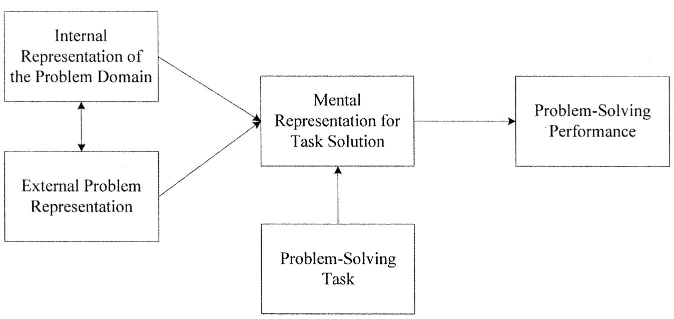

Fig. 4. Distributed model of cognitive fit.

Fig. 4 reflects the fact that in cognitive fit both the internal and external representation of the problem domain, and the interactions between them, as well as the problem-solving task, contribute to the development of the MENTAL REPRESENTATION FOR TASK SOLUTION. Rather than using the generic term, internal problem representation, we use the more specific term, INTERNAL REPRESENTATION OF THE PROBLEM DOMAIN. Note that this extended model of cognitive fit allows the researcher to consider independently the roles and nature of the internal representation of the problem domain, the EXTERNAL PROBLEM REPRESENTATION, and the mental representation for task solution, thereby distinguishing the internal representation from the mental representation in the prior model (Fig. 3).

Cognitive Fit in Geospatio–Temporal Schema Comprehension To apply the theory of cognitive fit to comprehension of the geospatio–temporal data semantics, we need to examine both the nature of the tasks and of the representations to be investigated. Because a number of aspects of fit are possible within the context of the theory of cognitive fit (see [45]), we first present the theoretical basis for the type of fit that we propose in the current context and then address fit in relation to the problem representation and the problem-solving task investigated here.

Propositional Representations: Prior research differentiates between analogical and propositional representation [46]. An analogical representation (for example, geographic maps) **depicts** a correspondence between the structure of the representation and the thing represented. A propositional representation, on the other hand, **describes** the thing represented. We focus on propositional representations in this paper.

Anderson and Bower's model of memory, known as the Human Associative Memory (HAM) model, is, in effect, a representation of the way that human problem solvers conceive of specific aspects of the tasks they are trying to solve [47]. We use HAM as presented in Anderson and Bower [48] for the purpose of developing theory relating to propositional representations (see Fig. 5). The nodes in the tree represent ideas, while the links represent relations or associations among ideas. Nodes in the proposition tree are represented with lower case letters and the labels on the arrows by upper case letters.

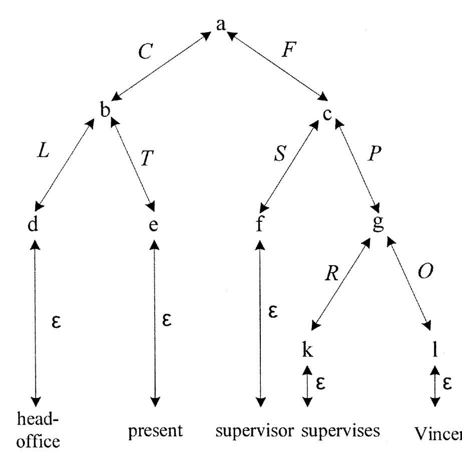

Fig. 5. Model of human associative memory adapted from Anderson and Bower [48].

In this model, each propositional tree (a) is divided into two subtrees: a fact subtree (c) and a context subtree (b); that is, propositions are composed of a (F)act and a (C)ontext. A context represents the situation in which the fact is true. Hence, node (a) represents the idea of the total proposition, node (b) the idea of the context, and node (c) the idea of the fact. Hence, the proposition asserts that fact (c) is true in context (b). A fact can be further subdivided into a (S)ubject (f) and a (P)redicate (g), and a predicate can be further subdivided into a (R)elation (k) and an (O)bject (l). And the context node (b) can be further subdivided into a (L)ocation node (d) and a (T)ime node (e).

Such a tree would be formed by parsing the sentence, *In the head-office, a supervisor supervises Vincent*, as shown in Fig. 5. This example shows clearly that the fact is formed by the association between the subject, in this case, *supervisor*, and the predicate, which in this case is made up of the relation *supervises* and the object *Vincent*. The context is provided by the location (*head-office*) and time nodes (*present* time). Of course, we have shown an example of a very simple propositional structure. In practice, a text could be composed of a number of propositions that correspond to multiple clauses in the sentence structure. Similarly, a propositional representation (for example, a conceptual schema) could be composed of a number of propositions that correspond to different parts of the schema.

External Problem Representations: External problem representations are the conceptual schemas that human problem solvers (in this case, data analysts) must comprehend. The problem representation that we address in this research is based on the entity-relationship diagram. Prior research (for example, [49]) has formally defined the data semantics of the ER Model using first order logic, which is a propositional representation.

As we have seen, the entity-relationship diagram can be augmented to represent geospatio–temporal data semantics. Such when/where data semantics can be represented both as a list that is separate from, and additional to, the schema, in a 1-LA representation and as annotations on the schema itself, in a 2-LA representation.

Perusal of our conceptual schemas and the HAM model of internal memory reveals the similarity between the 2-LA representation of the geospatio–temporal data semantics and HAM, and the dissimilarity between the 1-LA representation and HAM. Hence, the propositions of our 2-LA external problem representation are structurally analogous to the way in which propositions are stored in memory, as evidenced in HAM.

Problem-Solving Tasks: We first examine the types of problem-solving tasks investigated in this research, followed by the representations of the task that are developed by the problem solver.

As indicated above, a number of different types of schema understanding tasks have been investigated in conceptual modeling research. In this study, we examined two types of tasks: syntactic and semantic tasks. SYNTACTIC TASKS require an understanding of just the syntax (that is, the conceptual model) associated with a schema. For example, the syntax for an entity type is a rectangle. The comprehension validation tasks of Kim and March are examples of syntactic tasks [38]. SEMANTIC TASKS require an understanding of the meaning of the constructs in the schema. For example, an entity type represents a collection of objects, things, events, or places (in the real world), while a relationship represents an association between or among entities. The modeling correctness tasks of Batra, Hoffer, and Bostrom [50] and the discrepancy checking tasks of Kim and March [38] are examples of semantic tasks.

Our tasks are presented textually as English language questions in what we refer to as the EXTERNAL TASK REPRESENTATION. The propositions represented in these questions are then parsed according to the HAM model in such a way as to represent meaning in internal memory. Hence the propositions derived from these texts will also be represented internally according to the HAM model in what we call the INTERNAL TASK REPRESENTATION (see Fig. 6).

Hypotheses Fig. 6 presents the conceptual model of cognitive fit appropriate to this investigation of the relationship between the data analyst's internal representation of the task and the external problem representation that is used to support task solution. We can assess fit by examining the way in which propositions derived from comprehension questions are stored in memory and the way in which the conceptual modeling elements are represented in the conceptual schema. In other words, we can determine the external problem representation that matches the way in which the comprehension questions used in this research, are stored in internal memory (internal task representation). This situation then represents cognitive fit, with the resultant effect of maximizing accuracy and minimizing time, that is, problem-solving performance.

Clearly, a traditional conceptual schema (external problem representation) represents the what of a given application. From our prior discussion of the HAM model, it is clear that the *fact* part of the tree can represent quite well an entity (the **subject**) that has a particular **relationship** with another entity (the **object**). For example, in Fig. 8, the unary relationship *supervises* with *SALES_PERSON* represents a proposition: "a supervisor (sales person) supervises a subordinate (sales person)." This discussion illustrates, then, how the fact part of the proposition stored in internal memory matches the conceptual schema, which represents what is important to a particular application. Hence, the internal task representation (corresponding to a proposition such as "a supervisor salesperson supervises a subordinate salesperson"), as represented by HAM, matches the way in which information is conveyed in the external problem representation, that is, the conceptual schema. Note that the data semantics in a given application will be represented in the conceptual schema via a number of propositions.

The prior discussion of the HAM model (Fig. 5) revealed that the context part of the tree represents geospatio–temporal aspects of the statement, that is, **location** and **time**. The context conveys the information that a supervisor supervises a subordinate "in the head-office at the present time." Hence, the location and time of the proposition subtree reflect the geospatio–temporal data semantics that provide the context in which the fact is true, that is, the when/where. Note, then, that the conceptual schema (that is, the external problem representation) that uses a second level of abstraction to represent geospatio–temporal data semantics (2-LA) matches

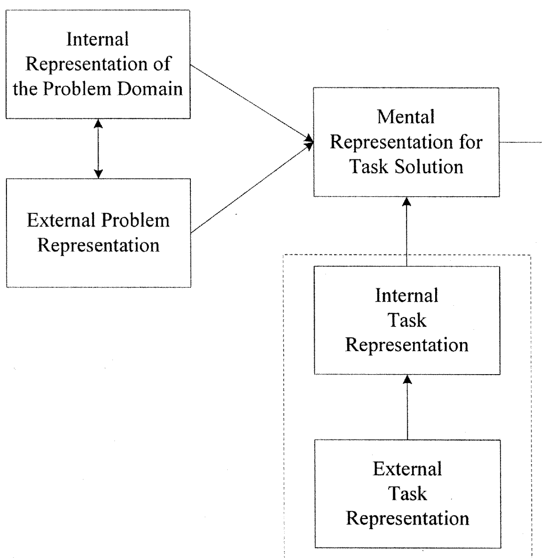

Fig. 6. Distributed cognitive fit model elaborating on task representation.

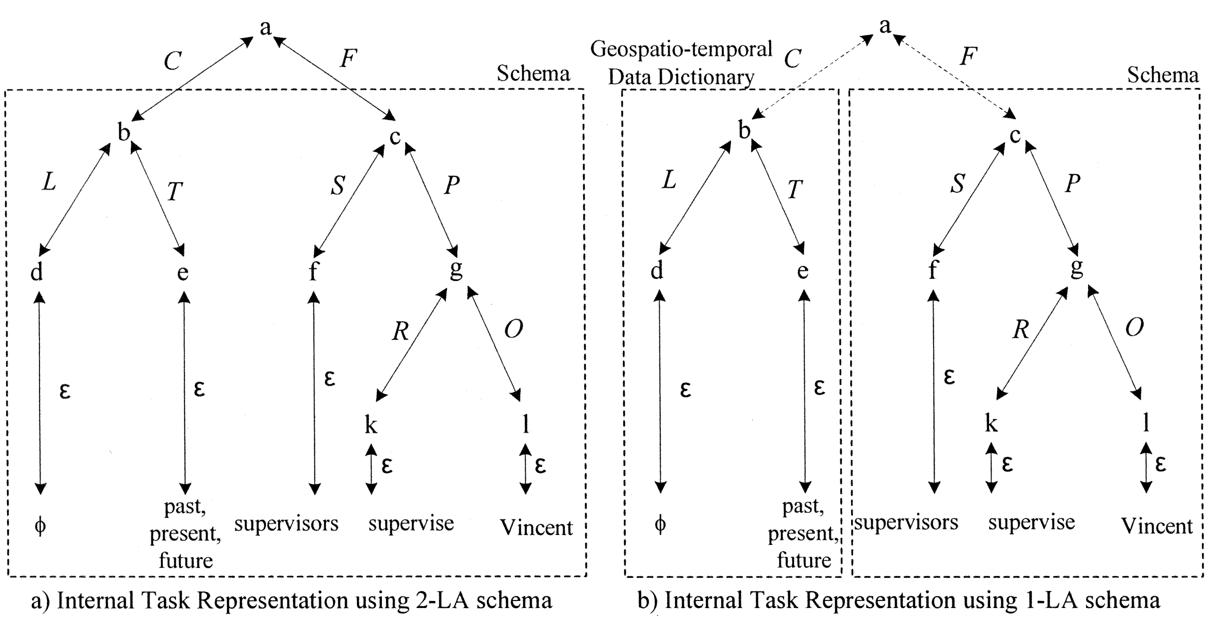

Fig. 7. Internal task representation using 2-LA and 1-LA schemas.

the way in which the task is represented in internal memory, which means that it matches the problem solver's internal task representation. Correspondingly, the separate representation of geospatio–temporal data semantics using 1-LA does not match the problem solver's internal task representation.

Fig. 7 below shows a problem solver's internal task representation of the following question: "On any given *day*, *Vincent*, a *SALES_PERSON*, is *supervised* by multiple *supervisor(s)*." The dotted boxes superimposed on the HAM model show how each of the representations addresses the problem of representing geospatio–temporal data semantics. In the 2-LA representation, the dotted box depicting the schema (external problem representation) in Fig. 7(a) accounts for both the fact subtree (node c) and the context subtree (node b). On the other hand, in the 1-LA representation shown in Fig. 7(b), the dotted box depicting the schema represents only the fact subtree, which specifies what (node c), and the dotted box depicting the geospatio–temporal data dictionary represents the context subtree, which specifies when and where. Hence the 2-LA representation matches the problem solver's internal task representation, while the 1-LA representation does not.

Hence, based on the theory of cognitive fit, we propose that problem solvers perform better with the 2-LA than with the 1-LA representation. Note that Siau refers to schemas that present the same information in different ways as informationally equivalent [51]. We therefore state the following proposition.

Proposition 1: Performance with an **external problem representation** that matches the problem solver's **internal representation of the task** is better than one that does not.

Table I summarizes the effects on problem-solving performance (accuracy and time) in the context of the external problem representation (1-LA and 2-LA) and internal task representation of syntactic and semantic tasks. Because our theory does not suggest differences in the comprehension of syntactic and semantic tasks, we expect that the effects of the

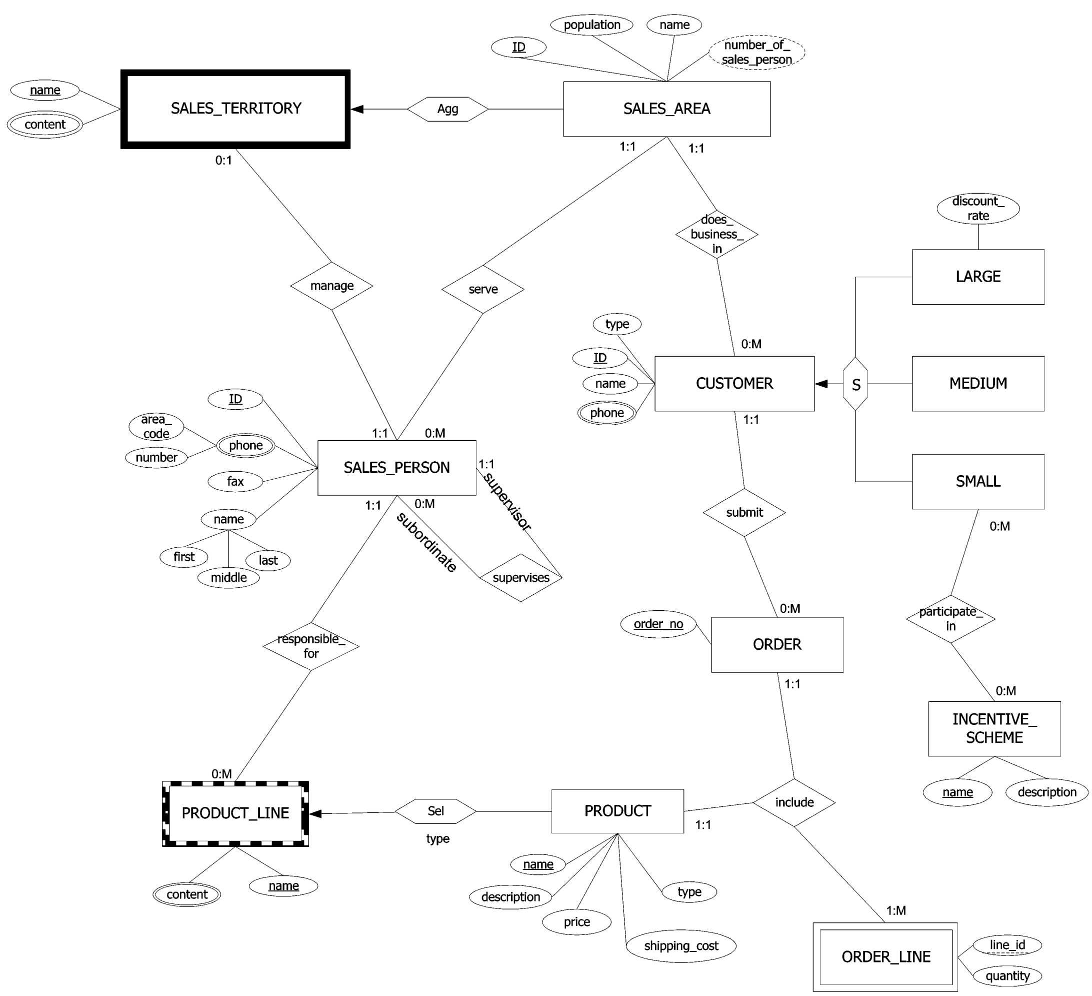

Fig. 8. Geospatio–temporal schema with annotations in the data dictionary.

different types of external problem representations will be similar for both types of tasks (see, also, [43]).

Specifically, we investigate the following hypotheses.

Hypothesis 1a: Problem solvers using a **2-LA problem representation** are more accurate on syntactic comprehension tasks than those using a **1-LA representation**.

Hypothesis 1b: Problem solvers using a **2-LA problem representation** are more accurate on semantic comprehension tasks than those using a **1-LA representation**.

Hypothesis 1c: Problem solvers using a **2-LA problem representation** are quicker on syntactic comprehension tasks than those using a **1-LA representation**.

Hypothesis 1d: Problem solvers using a **2-LA problem representation** are quicker on semantic comprehension tasks than those using a **1-LA representation**.

Prior research (see, for example, [50]) also suggests that the extent to which using a conceptual model will be free of mental effort can be evaluated via perceived ease of use. We therefore state the following proposition and corresponding hypothesis.

Proposition 2: An **external problem representation** that matches the problem solver's **internal representation of the task** is perceived to be easier to use than one that does not.

Hypothesis 2: Problem solvers perceive a **2-LA representation** to be easier to use than a **1-LA representation**.

TABLE I
SUMMARY OF EFFECTS OF 1-LA AND 2-LA REPRESENTATIONS

External Problem Representation	Analysis of Effects	Problem Solving Performance		Ease of Use
		Accuracy	Time	
1-LA	Indirect relationship between external problem representation and internal task representation • Cognitive fit does **not** exist • Further processing required	Lower	More	Lower
2-LA	Direct relationship between external problem representation and internal task representation • Cognitive fit exists • No further processing required	Higher	Less	Higher

TABLE II
GEOSPATIO–TEMPORAL DATA DICTIONARY FOR THE SALES APPLICATION

Abstraction	Temporal/Spatial Description	Granularity
SALES_PERSON	The lifespan of a sales person is specified as "state;" transaction time is not required	Day
SALES_PERSON	The position of sales person is specified as "point" on the horizontal surface	Degree
supervises	A "state" is associated with supervises; transaction time is not required	Day
PRODUCT.price	The history of the price of a product is specified as "state;" transaction time is not required	Day
PRODUCT.shipping_cost	The shipping cost for a product is region specific	Degree
CUSTOMER	The lifespan of a customer is specified as "state;" transaction time is not required	Day
CUSTOMER	The spatial position of a customer is specified as "point" on the horizontal surface	Degree
CUSTOMER.type	The history of the type of a customer is specified as "state;" transaction time is not required	Day
SALES_AREA	The lifespan of a sales area is specified as "state;" transaction time is not required	Day
SALES_AREA	The geometry of a sales area is specified as "region" on the horizontal surface	Degree
SALES_TERRITORY	The lifespan of a sales territory is specified as "state;" transaction time is not required	Day
SALES_TERRITORY	The geometry of a sales territory is specified as "region" on the horizontal surface	Degree
SALES_TERRITORY	The shape of a sales territory over the horizontal surface may change over time	Not Applicable
ORDER	The point of time at which the order is created is specified as "event;" transaction time is required	Minute

RESEARCH METHODOLOGY

We conducted a laboratory experiment to test our hypotheses.

Task Setting We investigated two types of annotation-based representations: 2-LA, in which geospatio–temporal aspects are represented on the schema, and 1-LA, in which geospatio–temporal aspects are represented in the geospatio–temporal data dictionary. We used a sales schema as the stimulus material because we expected that participants drawn from a business school would be familiar with a sales application.

Participants, who formed part of a database class, had been taught conventional conceptual modeling using the Unifying Semantic Model (USM) [8]. Before this experiment was conducted, the participants had already completed multiple conceptual modeling exercises in class, taken an exam that evaluated their conceptual modeling knowledge, and developed a conceptual schema for a real-world group project under the supervision of their instructor.

Participants The participants in this study were graduate students in a large public university in the southwest US. A total of 40 graduate students participated in the laboratory study. All the respondents were less than 45 years old with the majority being between 25 and 35 years. Respondents included almost an equal number of men (57.5%) and women (42.5%). Two-thirds of the respondents (67.5%) claimed to have some prior conceptual modeling experience. Almost everyone (95%) had some work experience, and many respondents (75%) had some database-related experience.

Cash prizes of $100, $50, and $25 were awarded to the top three participants, respectively, to motivate participants to perform as well as possible.

Experimental Design Participants were randomly assigned to two groups: 1-LA and 2-LA. Hence we used a 1×2 between-subjects design.

We conducted manipulation checks to assess the assignment of participants across treatments. There were no differences based on independent samples t-tests in any of the control variables (age, gender, prior education, major, and work experience) indicating that the participants were effectively randomized across treatments.

Experimental Materials Here we describe the conceptual schemas for both 2-LA and 1-LA, as well as the types of questions used to assess performance. Our training notes and the syntax formalism were reviewed by an independent database design expert to ensure that the material was not biased toward any group. Respondents in both groups were provided with notes related to the syntax formalism. Because prior research has observed that problem solvers may trade off accuracy for time, we measured both the accuracy and the time to complete the various tasks.

Schemas: Each participant was presented with one of the two semantically identical schemas: the 1-LA representation (see Fig. 8 and Table II) and the 2-LA representation (see Fig. 9). Both 1-LA and 2-LA participants were given the data dictionary presented in Appendix A. A 1-LA representation presents the geospatio–temporal annotations separately in the data dictionary (Table II), while 2-LA presents a geospatio–temporal schema with the annotations represented on the schema. For the 1-LA group, the schema and the geospatio–temporal data dictionary were presented on separate sheets of paper; the participants could place the schema and geospatio–temporal data dictionary side-by-side.

USM was used to develop the schema for the 1-LA group shown in Fig. 8. The schema was based on a sales application that included entity types such as *SALES_TERRITORY*, *SALES_AREA*, *CUSTOMER*, *ORDER*, *PRODUCT*, and *PRODUCT_LINE*. Each entity type had attributes represented in ovals. The schema also included supertype and subtypes: *SMALL*, *MEDIUM*, and *LARGE* were subtypes of the supertype *CUSTOMER*. The schema included a composite relationship (represented by a hexagon with *Sel*), a composite class *(PRODUCT_LINE)*, a grouping class (*SALES_TERRITORY*), and a grouping relationship (represented by a hexagon with *Agg*). Details related to these constructs are described elsewhere [8], [52]. The geospatio–temporal aspects associated with the schema were presented in the data dictionary shown in Table II. For example, the first two rows of the data dictionary show that the lifespan of the *SALES_PERSON* is represented as a state (i.e., a time period) with temporal granularity of *Day*, and that the position is represented as a point with geospatial granularity of *Degree*. Note that, because prior research suggests that there are few or no cues to navigation in graphical models [53], the items in the data dictionary are in random order; nor do they correspond with the schema in any way.

The geoSpatio–Temporal Unifying Semantic Model (ST USM) [14], [23], a geospatio–temporal conceptual model based on USM, was used to develop the 2-LA representation; the ST USM schema used in the experiment proper is shown in Fig. 9. As may be evident, the two representations (1-LA and 2-LA) were semantically identical and therefore informationally equivalent. For example, the annotation string (*S(day)/-//P(deg)/P(deg)/-*) associated with the *SALES_PERSON* implies that the lifespan of the *SALES_PERSON* is represented as state (*S*) with temporal granularity of *day* and that the geospatial position needs to be represented as point (*P*) with geospatial granularity of *deg*ree; additionally, "-" after "*S(day)/*" reflects the fact that transaction time is not required.

Schema Comprehension Tasks: Participants were presented with syntactic tasks followed by semantic tasks. As we have seen, syntactic tasks simply require an understanding of the syntax, while semantic tasks require an understanding of the meaning of the constructs in the schema. Responding to syntactic tasks requires observation of the conceptual schema/geospatio–temporal data dictionary, while responding to semantic tasks requires the problem solver to understand the meaning of the constructs in the schema/geospatio–temporal data dictionary.

The syntactic task was operationalized using 20 multiple-choice questions, which were similar to the comprehension validation tasks used by Kim and March [38]. Following is an example of a syntactic question: Name one simple temporal entity class whose lifespan is captured as state.

The semantic task consisted of 20 true/false questions, which were similar to the discrepancy checking tasks of Kim and March [38]. The participants were asked if the statements were consistent with the given schema. Following is an example of a semantic question: On any given *day*, *Vincent*, a *SALES_PERSON*, is *supervised* by multiple *supervisor(s)*. Note that different fonts were used to identify a concept from the schema (*Tahoma*) and

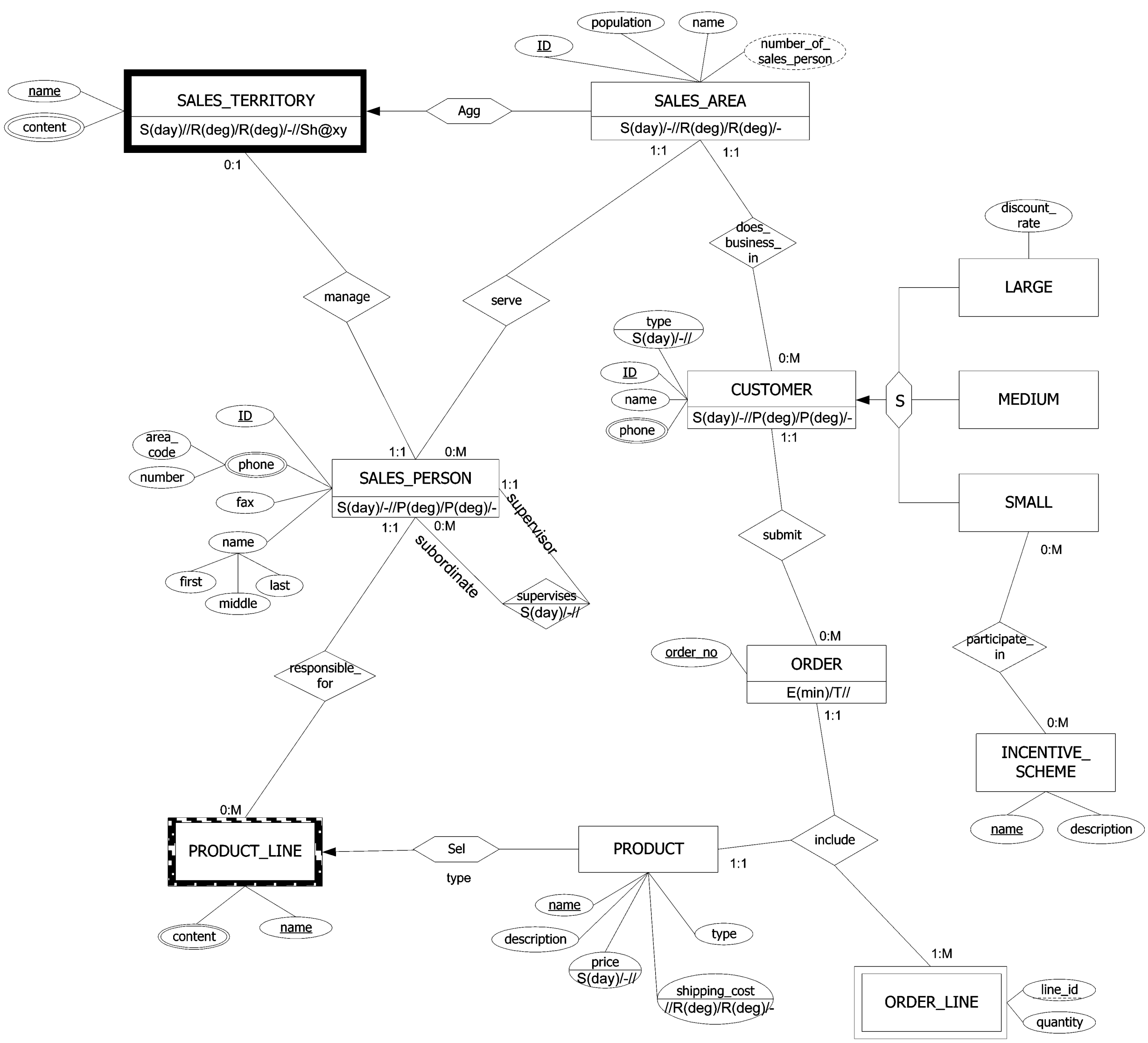

Fig. 9. Geospatio–temporal schema with annotations on the schema.

to identify instances associated with the concept (*Courier*). Appendix B presents the complete set of syntactic and semantic tasks used in the experiment.

Perceived Ease of Use: Perceived ease of use was operationalized as the degree to which an individual believes that using a particular system would be free of physical and mental effort [54]. The seven-item instrument used to measure perceived ease of use was adapted from Batra, Hoffer, and Bostrom [50].

Pilot Study We conducted a pilot study with graduate students who were engaged in database-related research. The pilot study helped us to eliminate ambiguity in the question wording, to test the experimental procedures, and to determine the length of time that the experiment would take to complete.

Experimental Procedure The participants first completed a brief demographic survey. Each session began with the first author providing an overview of around 40 minutes on conventional and geospatio–temporal conceptual modeling. Participants were then given two practice exercises, which took a total of 15 minutes. The practice session was followed by a debriefing of approximately 5 minutes. The experimental exercise, which took approximately 30 minutes, was then given to the participants. Following completion of the experimental tasks, participants completed a debriefing questionnaire that assessed their perceptions of the ease of use of the geospatio–temporal formalism that they used.

ANALYSIS AND RESULTS

In this section, we present the results of testing our hypotheses on performance (that is, accuracy normalized by maximum score as well as time) on both syntactic and semantic tasks using two representations of the geospatio–temporal data semantics, 2-LA and 1-LA. We also present our results related to perceptions of ease of use.

Table III presents the means, standard deviations (in parentheses), and statistical comparisons of normalized accuracy and time taken to complete the tasks. We report both statistical significance (p-values) and effect sizes (Cohen's d). Cohen's d, an effect size measure, is based on standardized group-mean differences and is pertinent where comparison of group means is of primary interest [55]. According to Cohen, a "small" effect size ($d = 0.2$) implies that 14.7% of the distributions of two populations are not overlapping, a "medium" effect size ($d = 0.5$) implies that 33% of the distributions of the two populations are not overlapping, and a "large" effect size ($d = 0.8$) implies that 47.4% of the distributions of the two populations are not overlapping [56].

An independent samples t-test suggested that the overall normalized accuracy for the 2-LA group was better than that for the 1-LA group ($t = 1.93$; $df = 38$; $p\text{-value} = 0.03^{**}$). Perusal of Table III indicates that participants using the 2-LA representation were more accurate on both syntactic and semantic tasks than the participants using the 1-LA representation. The size of the effect of different types of representations on comprehension performance is medium. Hence, Hypotheses 1a and 1b are supported. Table III also shows that the different types of representations have little effect on the time taken to complete the syntactic and semantic comprehension tasks. Hence, Hypotheses 1c and 1d are not supported. Further, the 2-LA representation was perceived to be significantly easier to use than the 1-LA representation, and the effect size was large, thus supporting Hypothesis 2.

TABLE III
BETWEEN-SUBJECT COMPARISON OF PERFORMANCE

	Facet	Max	2-LA group	1-LA group	p-value [a]	Cohen's d [b]
Accuracy [c]	**Syntactic**	**1**	**0.93 (0.08)**	**0.86 (0.15)**	**0.05****	**0.53****
	Semantic	**1**	**0.88 (0.09)**	**0.83 (0.10)**	**0.05****	**0.54****
Time	**Syntactic**	**NA**	**8.70 (3.71)**	**9.00 (1.86)**	**0.37**	**0.10***
	Semantic	**NA**	**9.50 (2.44)**	**9.80 (2.14)**	**0.34**	**0.13***
Perceptions	**Ease of Use** [d]	**7**	**2.17 (0.96)**	**2.98 (0.94)**	**<0.01*****	**0.85*****

a Statistical significance (p-value): *** significant at $p \le 0.01$; ** significant at $p \le 0.05$; * significant at $p \le 0.1$

b Effect size (Cohen's d): *** large (d=0.8); ** medium (d=0.5); * small (d=0.2)

c Note that the p-values for differences in accuracy performance on syntactic and semantic tasks for the 1-LA and 2-LA groups for the geospatio-temporal tasks alone were 0.04 and 0.19, respectively; see Appendix B for the specific tasks used in the experiment

Discussion and Implications

With recent advances in technologies such as high-resolution satellite-borne imaging systems, mobile systems, global positioning systems, and the overall decrease in hardware costs, temporal and geospatial data are finding their way into many traditional applications. Consequently, many DBMS vendors are incorporating capabilities to manage geospatial (for example, Oracle Spatial [57] and Informix Geodetic DataBlade [58]) and temporal (for example, Oracle Time Series Cartridge [59] and Informix TimeSeries DataBlade [60]) data. Recent research in geospatio–temporal conceptual modeling proposes augmenting conventional conceptual models with annotations. Yet we do not know how annotations, which represent additional data semantics **on** the schema, impact the understanding of the data analysts who interpret them. In this research, we examined the overall research question: How should geospatio–temporal data semantics be represented to effectively support data analysts in schema comprehension? To do so, we conducted a laboratory experiment to help understand the effects of augmenting a conventional conceptual model with geospatio–temporal annotations.

Below, we discuss the findings and the contributions of our research. We conclude with implications for both future research and practice.

Discussion of the Findings Our research resulted in three specific findings. First, we found that geospatio–temporal data semantics represented via a supplementary level of abstraction on the schema (2-LA) facilitated comprehension to a greater extent than the schema with one level of abstraction (1-LA). Second, we found no differences for time, suggesting that the effect of levels of abstraction is manifested in accuracy alone. Third, we found that 2-LA schemas are perceived to be easier to use than 1-LA schemas.

Our research makes a number of contributions to knowledge. First, it shows that the theory of cognitive fit serves as an appropriate theoretical basis for understanding how different geospatio–temporal data representations support schema comprehension. The competing perspective, that of the proximity compatibility principle, would also suggest that the 2-LA representation would prove superior to presenting when/where information in a geospatio–temporal data dictionary (1-LA) because of the "distance" of the 1-LA data from the base data presented on the schema. Note, however, that such a representation uses English language statements that are much simpler than the additional syntax required in 2-LA. Hence, to fully test whether the proximity compatibility principle can also serve as a theoretical basis for such research, an experiment would need to be conducted to compare the 1-LA representation with a representation in which the 2-LA space/time appears in the data dictionary. That is, performance can only be compared for syntactically identical annotation phrases in the two schemas.

Second, we viewed conceptual schema comprehension of geospatio–temporal data semantics in terms of matching the EXTERNAL PROBLEM REPRESENTATION to the INTERNAL TASK REPRESENTATION, based on theory on the structure of memory [47], [48] and of the storage of text in memory [61], [62]. Because prior research suggests that conceptual models for geographic data representation do not explicitly incorporate how humans cognitively store and use geographic data (see, for example, [18]), our conceptualization provides insights into the storage of geospatio–temporal data with consequent implications for comprehension of such conceptual models.

Third, this study suggests that annotations on the schema are appropriate for data analyst schema validation during conceptual design. While prior research [63] suggests that expressiveness needs to be balanced with simplicity and understandability, geospatio–temporal annotations on the schema can help achieve both expressiveness **and** understandability: they enhance the expressive power of conceptual models while preserving the level of schema comprehension. Thus, this study provides insights into how conventional conceptual models, such as the ER model, can be augmented for specialized applications, for example, the geospatio–temporal applications examined here.

In summary, then, while this paper presents preliminary findings, to the best of our knowledge, this is the first study to explore the representational effects of annotating a conceptual schema with geospatio–temporal annotations.

Our study has the following limitations. First, because we ran the experiment in nine experimental sessions, it is possible that there might have been differences across sessions. We took a series of measures to ensure equivalence of the sessions: (1) the same instructor presented the instructions in all sessions; (2) the same presentation was used consistently across the sessions; and (3) participants were allowed to ask only clarification questions. Second, we conducted our investigation with students who were relatively inexperienced in using real world conceptual modeling. Hence our findings do not generalize to professional conceptual designers who have a wealth of experience. Third, this study was conducted in a laboratory setting in which many aspects that might have come into play in more realistic settings were controlled. Fourth, in our experiment, we did not differentiate between the syntactic and semantic tasks we examined because our theory suggested that there should be no differences.

Implications of the Findings Here we examine the implications of our findings for future research and practice.

Implications for Research: Our research has a number of implications for researchers. First, research needs to be conducted to establish the boundaries of our theory. For example, because our findings are generalizable only to novice data analysts, future research should be undertaken to examine the applicability of our findings to the general population of IS professionals who have years of conceptual modeling experience. Research also needs to be conducted to determine whether our theory might serve as a foundation for other types of geospatio–temporal conceptual models with 2-LA representations, for example, MADS [32], which uses pictogram-oriented geospatio–temporal representations.

Second, we believe that researchers could benefit from using our enhanced model of cognitive fit in their studies. The HAM model is very effective in representing tasks that are presented as text. And, again, the structure of HAM is structurally similar to the what and when/where data semantics that can be represented in a conceptual schema.

Third, future research should investigate the effects of different types of external problem representations (that is, 1-LA and 2-LA) on tasks of greater complexity, for example, conceptual schema understanding tasks that involve problem-solving; see, for example, [43].

Fourth, future research should also differentiate the findings based on geospatio–temporal and nongeospatio–temporal tasks.

Fifth, conceptual schemas, including geospatio–temporal schemas, are an important vehicle for the technical communications that occur during information systems development. Siau and Tan suggest that using conceptual schemas for this purpose can lead to shared understanding among various stakeholders, such as users, managers, and developers [64]. Further research should investigate the effect of annotated schemas on technical communicators who play an important role in reducing maintenance costs and programming time, lowering support and training costs, and helping reduce user errors [65].

Finally, we speculate, based on our observations in this research, that problem representations need to represent conceptual elements distinctly but in an integrated fashion. In our research, the 2-LA schema representation presented the two aspects, what and when/where, as distinct elements integrated into the same diagram. A representation with a supplementary level of abstraction provides a mechanism that presents the when/where data semantics orthogonal to the what data semantics while keeping the two aspects tightly coupled. Future research could investigate the reasons underlying the improved performance with the 2-LA representation, for example by conducting protocol analysis studies, which are particularly effective at examining problem-solving processes.

Implications for Practice: Our research has several implications for practice. First, we expect that database analysts in practice would have little difficulty understanding both the syntax and the semantics of the annotated schema because the annotations are built on top of traditional conceptual schema. This implies that if the 2-LA approach were to be adopted by data analysts, we would expect that training time related to representing the geospatio–temporal aspects would not be substantial and that the training costs for organizations would therefore be minimal.

Second, the 2-LA approach could be used as the basis for augmenting existing CASE tools. Such an approach would be straightforward to implement [66] and, as we have seen above, straightforward to comprehend.

CONCLUSIONS

Many real-world geospatial and temporal applications need to organize data based on time and/or geographic space. Recent research in geospatio–temporal conceptual modeling, which advocates representing geospatio–temporal data semantics on the schema, challenges the prevalent practice of representing the geospatio–temporal data semantics in a data dictionary outside the schema. In this study, we compared the representational effects of geospatio–temporal annotations on the schema and those presented in a data dictionary.

We examined conceptual schema comprehension of geospatio–temporal data semantics from the viewpoint of the cognitive fit of the external problem representation to the internal representation of the comprehension task (internal task representation). Our foundational theory, the theory of cognitive fit, suggests that an external problem representation that matches the problem solver's internal task representation enhances performance. We used theories relating to the representation of meaning in memory to understand the mental representation the problem solver forms on interacting with the external problem representation and theories relating to the representation of text in memory as the basis for our understanding of the structure of the internal task representation.

We conducted a laboratory experiment to assess performance on geospatio–temporal schema comprehension tasks supported by two semantically identical schemas (external problem representations); one of the schemas mapped closely to the internal task representation, while the other did not. We found that the geospatio–temporal conceptual schemas that corresponded to the internal representation of the task enhanced the accuracy of schema comprehension; comprehension time was equivalent for both. Thus, our findings suggest that the cognitive fit between the internal task representation and the conceptual schemas representing the geospatio–temporal data semantics in the form of annotations was manifested in accuracy of schema comprehension and not in time for problem solution.

APPENDIX A

TABLE IV
DATA DICTIONARY FOR THE SALES APPLICATION

Entity Class	Attribute	Description
SALES_PERSON	ID	Company allocated random number that identifies a sales person
	name	The name of a sales person
	first	The first name of a sales person
	middle	The middle name of a sales person
	last	The last name of a sales person
	fax	The fax number of a sales person
	phone	The phone number of a sales person
	area_code	The area code of a sales person's phone; e.g., (520)
	number	The number of a sales person's phone; e.g., 621-2748
SALES_TERRITORY	ID	ID for a sales territory
	population	The number of people who live in a sales territory
	name	The name of a sales territory
	number_of_sales_person	The number of sales persons who work in the sales territory
SALES_DIVISION	name	The name of a sales region
	content	The sales territories in the sales region
CUSTOMER	ID	ID for a customer
	name	The name of a customer
	phone	The phone number where a customer can be reached
	type	The type of a customer, i.e., small, medium or large
LARGE	discount_rate	A flat discount rate (in percentage) that is offered on all orders to a large customer
ORDER	order_no	Order number for an order
PRODUCT	name	The name of a product
	price	The price of a product
	description	The description of a product
	shipping_cost	The shipping cost for a product
PRODUCT_LINE	content	The products in a product line
	name	The name of a product line
INCENTIVE_SCHEME	name	The name of an incentive scheme
	description	The description of an incentive scheme; e.g., a trip to Hawaii, if the sales are larger than $100,000

APPENDIX B

TABLE V
EXPERIMENTAL TASKS

Syntactic Tasks	Semantic Tasks
Based on the schema/data dictionary, please answer the following questions.	Indicate whether the statements below are true/false with respect to the schema/data dictionary.
Non Geospatio-Temporal Tasks	
• Name one *ternary interaction relationship*. • Name one *multi-valued attribute* of "SALES_PERSON". • Name one *single-valued composite attribute* (specify the entity class also). • Name one *subclass*. • Name one *composite class*. • Name one *superclass*. • Name one *unary relationship*. • What are the *total number* of *subclasses* in the schema. • Name one *weak entity class*. • Name one *identifying attribute* of "ORDER_LINE".	• 2625XE modem, a "PRODUCT," has to have an associated "description." • ARQ Inc., a "LARGE CUSTOMER," must have an "ID" and is offered a "discount_rate." • John, a "SALES_PERSON," can have three "phone" numbers. • Jennifer, a "LARGE CUSTOMER," can have no more than one "phone" number. • Ethernet Card, a "PRODUCT_LINE," includes the same "type" of "PRODUCT"(s); e.g., the "type" of the "PRODUCT"(s) for Ethernet Card is EC. • Flat Screen Monitor, a "PRODUCT_LINE," may optionally have one "SALES_PERSON" who is "responsible_for" this "PRODUCT_LINE." • Jennifer, a "CUSTOMER," "does_business_in" exactly one "SALES_AREA." • John, a "SALES_PERSON," "supervises" other "SALES_PERSON" like Tony, Joan, and Michelle. • AZ_256, a "SALES_TERRITORY," may be composed of multiple "SALES_AREA." • Jacob, a "SMALL" "CUSTOMER," can optionally "participate_in" at most one "INCENTIVE_SCHEME."
Geospatio-Temporal Tasks	
• Name one *simple temporal entity class* whose lifespan is captured as state. • Name one *time-varying spatial grouping class*. • Name one *simple spatial entity class* whose geometry is specified as *region* on the surface of the Earth. • Name one *bitemporal* (having both valid time and transaction time) *entity class*. • Name one *temporal attribute* (specify the entity class also). • Name one *temporal relationship*. • Name one *simple spatial entity class* specified as a *point* on the surface of the Earth. • What are the *total number* of *temporal attributes* in the schema. • Name one *spatial attribute* (specify the entity class also). • Name one *spatial relationship*.	• On any given "day," Vincent, a "SALES_PERSON," is "supervise" by multiple "supervisor." • Jonathan, a "SALES_PERSON," has an associated lifespan of existence, from the recruitment "day" to the resignation "day," • The spatial extent of AZ_230, a "SALES_AREA"—represented as region—may change; the spatial extent (i.e., "shape" of regions) of, e.g., AZ_230, on different "day" is captured. • The time at which ORD2357, an "ORDER," is stored in the database does not need to be captured. • The location of Jennifer, a "CUSTOMER," is represented by a point, the *x*-coordinate and the *y*-coordinate in "degree," on the surface of the Earth. • Flat Screen Monitor, a "PRODUCT_LINE," have been introduced only in some regions, e.g., Arizona and California. • The "shipping_cost" for a "PRODUCT," 2345xp, is not the same for various regions. • All the "supervisor" of Sarah during her existence as a "SALES_PERSON" are captured. • The time when "price" for MODEM2357B, a "PRODUCT," was stored in the database is captured. • The "price" of Monitor21, a "PRODUCT," is different for different regions.

ACKNOWLEDGMENT

This work was supported in part by NASA under Grant 314401. The authors are grateful to R. T. Snodgrass and S. P. Weisband for their help in conducting the experiment as well as to the editor, guest editors, H. C. Chan and K. Siau, and anonymous reviewers for their insightful and constructive comments.

REFERENCES

[1] G. Albaredes, "A new approach: User oriented GIS," in *Eur. Conf. Exhibition on Geographic Information Systems (EGIS '92)*, Munich, Germany, 1992, pp. 830–837.
[2] B. Kuipers, "Modeling spatial knowledge," *Cognitive Sci.*, vol. 2, no. 2, pp. 129–153, 1978.

[3] B. E. Mennecke and M. D. Crossland, “Geographic information systems: Applications and research opportunities for information systems researchers,” in *29th Hawaii Int. Conf. System Sciences*, 1996, pp. 537–546.
[4] C. Harder, *ArcView: GIS Means Business*. Redlands, CA: Environmental Syst. Res. Inst., Inc., 1997.
[5] M. F. Worboys, H. M. Hearnshaw, and D. J. Maguire, “Object-oriented data modeling for spatial databases,” *Int. J. Geograph. Inf. Syst.*, vol. 4, no. 4, pp. 369–383, 1990.
[6] P. P. Chen, “The entity-relationship model—Toward a unified view of data,” *ACM Trans. Database Syst.*, vol. 1, no. 1, pp. 9–36, 1976.
[7] R. Elmasri and S. B. Navathe, *Fundamentals of Database Systems*, 2nd ed. Redwood City, CA: Benjamin Cummings, 1994.
[8] S. Ram, “Intelligent database design using the unifying semantic model,” *Inf. Manage.*, vol. 29, no. 4, pp. 191–206, 1995.
[9] V. Khatri, S. Ram, R. T. Snodgrass, and I. Vessey, “Strong vs. Weak Approaches to Conceptual Design: The Case of Temporal Data Semantics,” Kelley School Bus., Bloomington, IN, TR 147-1, 2006.
[10] H. Gregersen and C. S. Jensen, “Temporal entity-relationship models—A survey,” *IEEE Trans. Knowl. Data Eng.*, vol. 11, no. 3, pp. 464–497, May/Jun. 1999.
[11] C. Parent, S. Spaccapietra, E. Zimányi, P. Donini, C. Plazanet, and C. Vangenot, “Modeling spatial data in the MADS conceptual model,” in *8th Int. Symp. Spatial Data Handling (SDH’98)*, Vancouver, Canada, 1998, pp. 138–150.
[12] N. Tryfona and C. S. Jensen, “Conceptual data modeling for spatiotemporal applications,” *Geoinformatica*, vol. 3, no. 3, pp. 245–268, 1999.
[13] P. Shoval and I. Frumermann, “OO and EER conceptual schemas: A comparison of user comprehension,” *J. Database Manage.*, vol. 5, no. 4, pp. 28–38, 1994.
[14] V. Khatri, S. Ram, and R. T. Snodgrass, “Augmenting a conceptual model with geospatiotemporal annotations,” *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1324–1338, Nov. 2004.
[15] C. D. Wickens and C. M. Carswell, “The proximity compatibility principle: Its psychological foundation and relevance to display design,” *Human Factors*, vol. 37, no. 3, pp. 473–494, 1995.
[16] I. Vessey and R. Glass, “Strong vs. weak approaches to systems development,” *Commun. ACM*, vol. 41, no. 4, pp. 99–102, 1998.
[17] A. Newell, “Heuristic programming: Ill-Structured problems,” in *Progress in Operations Research*, J. Aronofsky, Ed. New York: Wiley, 1969.
[18] J. L. Mennis, D. J. Peuquet, and L. Qian, “A conceptual framework for incorporating cognitive principles into geographical database representation,” *Int. J. Geographic Inf. Sci.*, vol. 14, no. 6, pp. 501–520, 2000.
[19] Y. Wand, D. E. Monarchi, J. Parsons, and C. C. Woo, “Theoretical foundations for conceptual modeling in information systems development,” *Decision Support Syst.*, vol. 15, no. 4, pp. 285–304, 1995.
[20] C. S. Jensen, C. E. Dyreson, M. Bohlen, J. Clifford, R. Elmasri, S. K. Gadia, F. Grandi, P. Hayes, S. Jajodia, W. Kafer, N. Kline, N. Lorentzos, Y. Mitsopoulos, A. Montanari, D. Nonen, E. Peresi, B. Pernici, J. F. Roddick, N. L. Sarda, M. R. Scalas, A. Segev, R. T. Snodgrass, M. D. Soo, A. Tansel, R. Tiberio, and G. Wiederhold, “A consensus glossary of temporal database concepts—February 1998 version,” in *Temporal Databases: Research and Practice*, O. Etzion, S. Jajodia, and S. Sripada, Eds. New York: Springer-Verlag, 1998.
[21] R. T. Snodgrass and I. Ahn, “Temporal databases,” *IEEE Computer*, vol. 19, no. 9, pp. 35–42, Sep. 1986.
[22] H. Gregersen and C. Jensen, “Conceptual Modeling of Time-Varying Information,” TIMECENTER, Tech. Rep. TR-35, Sep. 10, 1998.
[23] V. Khatri, S. Ram, and R. T. Snodgrass, “Supporting user-defined granularities and indeterminacy in a spatiotemporal conceptual model,” *Ann. Math. Artificial Intelligence*, vol. 36, no. 1–2, pp. 195–232, 2002.
[24] B. David, M. V. D. Herrewegen, and F. Salge, “Conceptual models for geometry and quality of geographic information,” in *Geographic Objects With Indeterminate Boundaries*, P. A. Burrough and A. Frank, Eds. New York: Taylor & Francis, 1996, p. 352.
[25] L. H. Hermosilla, “A unified approach for developing a temporal GIS with database and reasoning capabilities,” in *5th Eur. Conf. Exhibition Geographic Information Systems (EGIS ’94)*, 1994, pp. 122–131.
[26] S. Shekhar, M. Coyle, B. Goyal, D.-R. Liu, and S. Sarkar, “Data models in geographic information systems,” *Commun. ACM*, vol. 40, no. 4, pp. 103–111, 1997.
[27] A. Y. Tang, T. M. Adams, and E. L. Usery, “A spatial data model design for feature-based geographical information systems,” *Int. J. Geograph. Inf. Syst.*, vol. 10, no. 5, pp. 643–659, 1996.
[28] R. Elmasri, G. Wuu, and V. Kouramajian, “A temporal model and query language for EER databases,” in *Temporal Databases: Theory, Design and Implementation*, A. Tansel, Ed. Redwood City, CA: Benjamin Cummings, 1993, pp. 212–229.
[29] R. Elmasri and G. T. J. Wuu, “The temporal model and query language for ER databases,” in *6th Int. Conf. Data Eng.*, 1990, pp. 76–83.
[30] S. Ferg, “Modeling the time dimension in an entity-relationship diagram,” in *4th Int. Conf. Entity-Relationship Approach*, 1985, pp. 280–286.
[31] R. T. Snodgrass, *Developing Time-Oriented Database Applications in SQL* San Francisco, CA, 1999, Morgan Kaufmann Series in Data Management Systems.
[32] C. Parent, S. Spaccapietra, and E. Zimanyi, “Spatio-temporal conceptual models: Data structures + space + time,” in *7th ACM Symp. Advances in Geographic Information Systems*, Kansas City, KS, 1999, p. .
[33] Y. Wand and R. Weber, “On the ontological expressiveness of information systems analysis and design grammars,” *Inf. Syst. J.*, vol. 3, no. 4, pp. 217–237, 1993.

[34] J. Zhang, "External representations in complex information processing tasks," in *Encyclopedia of Microcomputers*, A. Kent, J. G. Williams, and C. M. Hall, Eds. New York: Marcel Dekker, 2001, vol. 68, pp. 164–180.
[35] K. Siau, Y. Wand, and I. Benbasat, "The relative importance of structural constraints and surface semantics in information modeling," *Inf. Syst.*, vol. 22, no. 2–3, pp. 155–170, 1997.
[36] F. Bodart, A. Patel, M. Sim, and R. Weber, "Should optional properties be used in conceptual modeling? A theory and three empirical tests," *Inf. Syst. Res.*, vol. 12, no. 4, pp. 384–405, 2001.
[37] J.-M. Burkhardt, F. Détienne, and S. Wiedenbeck, "Object-oriented program comprehension: Effect of expertise, task and phase," *Empirical Software Eng.*, vol. 7, no. 2, pp. 115–156, 2002.
[38] Y.-G. Kim and S. T. March, "Comparing data modeling formalisms," *Commun. ACM*, vol. 38, no. 6, pp. 103–115, 1995.
[39] A. Burton-Jones and R. Weber, "Understanding relationships with attributes in entity-relationship diagrams," in *Proc. 20th Int. Conf. Information Systems*, Charlotte, NC, 1999, pp. 214–228.
[40] A. Gemino and Y. Wand, "Evaluating modeling techniques based on models of learning," *Commun. ACM*, vol. 46, no. 10, pp. 79–84, 2003.
[41] I. Vessey, "Cognitive fit: A theory-based analysis of graphs vs. tables literature," *Decision Sci.*, vol. 22, no. 2, pp. 219–240, 1991.
[42] J. Zhang and D. A. Norman, "Representations in distributed cognitive tasks," *Cognitive Sci.*, vol. 18, no. 1, pp. 87–122, 1994.
[43] V. Khatri, I. Vessey, V. Ramesh, P. Clay, and S.-J. Park, "Understanding conceptual schemas: Exploring the role of application and IS domain knowledge," *Inf. Syst. Res.*, vol. 17, no. 1, pp. 81–99, Mar. 2006.
[44] T. Shaft and I. Vessey, "The role of cognitive fit in the relationship between software comprehension and modification," *MIS Quart.*, vol. 30, no. 1, pp. 29–55, Mar. 2006.
[45] I. Vessey, "The theory of cognitive fit: One aspect of a general theory of problem solving?," in *Human-Computer Interaction and Management Information Systems: Foundations, Advances in Management Information Systems Series*, P. Zhang and D. Galletta, Eds. Armonk, NY: M.E. Sharpe, 2006.
[46] Z. Kulpa, "Diagrammatic representation and reasoning," *Machine Graphics and Vision*, vol. 3, no. 1/2, pp. 77–103, 1994.
[47] J. R. Anderson and G. H. Bower, *Human Associative Memory*. Washington, D.C.: V. H. Winston & Sons, 1973.
[48] ——, *Human Associative Memory: A Brief Edition*. Hillsdale, NJ: Lawrence Erlbaum, 1980.
[49] M. Gogolla and U. Hohenstein, "Toward a semantic view of an extended entity-relationship model," *ACM Trans. Database Syst.*, vol. 16, no. 3, pp. 369–416, 1991.
[50] D. Batra, J. A. Hoffer, and R. P. Bostrom, "Comparing representations with relational and EER models," *Commun. ACM*, vol. 33, no. 2, pp. 126–139, 1990.
[51] K. Siau, "Information and computational equivalence in comparing information modeling methods," *J. Database Manage.*, vol. 15, no. 1, pp. 73–86, 2004.
[52] S. Ram and V. C. Storey, "Composite and grouping: Extending the realm of semantic modeling," in *26th Hawaii Int. Conf. System Sciences (HICSS)*, 1993, pp. 80–89.
[53] B. C. Hungerford, A. R. Hevner, and R. W. Collins, "Reviewing software diagrams: A cognitive study," *IEEE Trans. Softw. Eng.*, vol. 30, no. 2, pp. 82–96, Feb. 2004.
[54] F. D. Davis, "A technology acceptance model for empirically testing new user end-user information systems: Theory and results," in *Sloan School of Management*. Boston, MA: MIT, 1985.
[55] X. Fan, "Statistical significance and effect size in educational research: Two sides of a coin," *J. Educat. Res.*, vol. 5, pp. 275–282, May/Jun. 2001.
[56] J. Cohen, *Statistical Power Analysis of the Behavioral Sciences*. New York: Academic, 1988.
[57] C. Murray, *Oracle Spatial User's Guide and Reference*: Oracle, 2000, vol. 2001.
[58] C. Scannell, D. Ashkenas, T. Day, R. Gutman, A. Guttman, J. Klebanoff, M. McDevitt, G. Morgan, C. Renck, R. Uleman, M. V. Lunteren, and J. Zhang, *Informix Geodetic Datablade Module: User's Guide*, 2001.
[59] "Oracle, "Oracle8™ Time Series Cartridge," Oracle Corp., Redwood City, CA, A57501-01, 8.0.4 ed., 1997.
[60] J. Simmonds, *Informix TimeSeries Datablade Module: User's Guide*, 2000.
[61] W. Kintsch, *The Representation of Meaning in Memory*. Hillsdale, NJ: Lawrence Earlbaum, 1974.
[62] D. Rumelhart, P. Lindsay, and D. Norman, "A process model for long-term memory," in *Organization of Memory*, E. Tulving and W. Donaldson, Eds. New York: Academic, 1972.
[63] C. Batini, S. Ceri, and S. B. Navathe, *Conceptual Database Design: An Entity-Relationship Approach*. Redwood City, CA: Benjamin Cummings, 1992.
[64] K. Siau and X. Tan, "Technical communication in information systems development: The use of cognitive mapping," *IEEE Trans. Prof. Commun.*, vol. 48, no. 3, pp. 269–284, Sep. 2005.
[65] J. Redish, "Adding value as a professional technical communicator," *Tech. Commun.*, vol. 42, no. 1, pp. 26–39, 1995.
[66] V. Khatri, S. Ram, and R. T. Snodgrass, "On augmenting database design-support environments to capture the geo-spatio-temporal data semantics," *Inf. Syst.*, vol. 31, no. 2, pp. 98–133, 2006.

Vijay Khatri (S'97–A'02–M'04) received the Ph.D. degree in Management Information Systems from the University of Arizona. He is an Assistant Professor of Information Systems in the Kelley School of Business, Indiana University, Bloomington, IN. His research interests center on developing techniques to facilitate elicitation of spatial/temporal database requirements, understanding the cognitive aspects that influence analysis and design of information systems, and examining the diffusion of information technology as well as user acceptance of new technologies. His work has been published or is forthcoming in *Information Systems Research*, *Information Systems*, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, *Annals of Mathematics and Artificial Intelligence*, and IEEE TRANSACTIONS ON EDUCATION, among others. He is a member of the Association of Information Systems (AIS), ACM, and Informs.

Iris Vessey received the M.Sc., MBA, and Ph.D. degrees in Management Information Systems from the University of Queensland. She is Honorary Professor at the University of Queensland and Adjunct Professor at the Queensland University of Technology, Brisbane, Australia. Her research interests focus on the evaluation of emerging information technologies, knowledge management systems, and the management and organization of enterprise resource planning systems (ERPs). She serves, or has served, as an Associate Editor at *Information Systems Research*, *Journal of Database Management*, *Journal of Management Information Systems*, *MIS Quarterly*, and *Management Science*, and serves on the Executive Board of Information Systems Frontiers. During the first eight years of its life, she served as Secretary of the Association for Information Systems (AIS), as well as of the International Conference on Information Systems (ICIS) following its merger with AIS. She is an inaugural AIS Fellow of the AIS.

Sudha Ram (S'85–M'85) received the B.S. degree in mathematics, physics, and chemistry from the University of Madras, in 1979, the PGDM degree from the Indian Institute of Management, Calcutta, in 1981, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, in 1985. She is Eller Professor, Management Information Systems in the Eller College of Management, University of Arizona, Tucson, AZ. She has published articles in such journals as *Communications of the ACM*, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, *Information Systems*, *Information Systems Research*, *Management Science*, and *MIS Quarterly*. She is a member of ACM, IEEE Computer Society, INFORMS, and the Association for Information Systems (AIS). She is also the director of the Advanced Database Research Group based at the University of Arizona.

V. Ramesh is an Associate Professor in the Information Systems Department and the Ford Motor Company Teaching Fellow at Indiana University's Kelley School of Business, Bloomington, IN. He is the director of the MS in Information Systems program. His research interests are in data modeling, heterogeneous databases, virtual teams and groupware, usability and software engineering. His research has appeared in leading journals, such as *MIS Quarterly*, *ACM Transactions on Information Systems*, *Communications of the ACM*, *Journal of Management Information Systems*, *Information Systems*, IEEE EXPERT, among others.